基于 Smooth-DETR 的产品表面小尺寸缺陷检测算法 *

张乃雪，钟羽中†，赵 涛，佃松宜

(四川大学 电气工程学院，成都 610065)

摘 要：为应对实际工业产品视觉质量检测中缺陷罕见、尺寸小等挑战，提出了一种仅需要少量训练样本的小尺寸缺陷检测算法——Smooth-DETR。该算法采用基于 DETR 的编码-解码结构对缺陷类别和位置进行预测，该结构降低了参数量和计算复杂度。因 DETR 强大的全局特征学习能力，该算法可从少量训练样本中充分挖掘产品表面纹理特征，从而对打破了表面纹理连续性的缺陷检出率高。通过结合 Smooth-L1 损失和 GIoU 损失的优势，进一步提升了小尺寸缺陷的回归精度。实验结果表明，所提方法检测性能优于现有先进检测模型。此外，仅用少量训练样本，该算法对 11 类产品表面的缺陷检测平均精确率就能够达到 98%以上。

关键词：Transformer；DETR 模型；GIoU 损失；表面缺陷检测；深度学习

中图分类号：TP391.4 doi: 10.19734/j.issn.1001-3695.2022.01.0011

Detection method for small-size defects based on Smooth-DETR

Zhang Naixue, Zhong Yuzhong†, Zhao Tao, Dian Songyi

(*College of Electrical Engineering*, *Sichuan University*, *Chengdu* 610065, *China*)

Abstract: To deal with challenges of limited and small-size defects in product quality inspection, this paper proposed a method for surface-defect-detection of small-size with few training samples (Smooth-DETR) . This method utilized DETR-based encoder-decoder to predict the classification and location of defects, which reduced the parameters and complexity. DETR has a strong global feature learning capability, which could obtain rich texture features of product surfaces with few samples, so that it is easy to detect defects that break the continuity of texture. The combination of Smooth-L1 loss and GIoU loss improves the regression accuracy on small-size defect samples. Experimental results show that the proposed method performs better than the existing state-of-the-art methods. Moreover, the average detection precision of the proposed method for 11 different classes of surface defects is higher than 98% with few training samples.

Key words: Transformer; DETR; GIoU loss; surface defect detection; deep learning

0 引言

工业产品质量检测是生产和运维中的重要环节，是保证产品外观、质量和性能的关键。近年来，随着深度学习的发展及其在各个场景的广泛应用[1, 2]，基于深度学习的目标检测算法也逐渐应用于表面缺陷检测任务中[3, 4]以提高生产和检测效率，同时降低人的主观意识对检测结果的影响。与传统的缺陷检测相比，基于深度学习的表面缺陷检测算法更加智能化，且具有更好的检测准确率和泛化性[5]。

基于深度学习的表面缺陷检测任务可以大致分为基于检测的表面缺陷检测算法和基于分割的表面缺陷检测算法。基于检测的表面缺陷检测通常可以被视作一个特殊的目标检测任务。目前目标检测技术以端到端的深度学习检测算法为主，分为两大类，一类是以 Faster R-CNN 为代表的 two-stage 检测算法[3]；另一类是以 YOLO 为代表的 one-stage 检测算法[6]。Di 等人[7]首先使用多组 CNN 对图像进行分类，针对不同类型的缺陷分别训练不同的卷积核；然后，将可能包含缺陷的特征映射输入到另一个基于 YOLO 的网络中，对缺陷的检测框进行回归。为应对 YOLOv4 的主干网络复杂度较高和难以检测到小尺寸表面缺陷等问题，Lian 等人[8]提出了 YOLOv4-DefectSP 算法。该算法利用深度可分离卷积代替传统的卷积结构，并融合知识蒸馏[9]，不仅提升了对小尺寸缺陷检测准确率，还极大地降低了模型复杂度。此外，为了提高织物数据集检测的速度和准确性，Zhou 等人[10]提出了更高效的 RCNN 算法。该算法利用可变形卷积网络代替了 Faster R-CNN 中主干网络的最后一个阶段，并引入特征金字塔网络和距离 IoU 损失函数，以达到检测更加快速和准确的目的。除了基于目标检测的缺陷检测方法以外，也有不少其他的检测方法。例如，由于很多行业无法获取足够的高精度标注样本，文献[11]提出了一个混合监督网络，由分割网络和分类网络两个子网络构成。该网络利用图像级标签和区域级标签混合监督的方法达到缺陷检测的目的，在一定程度上减少了对标注样本的需求。另外，基于分割的表面缺陷检测是指将语义分割模型用于解决表面缺陷检测任务。通常，语义分割模型由编码-解码结构组成，常见的有 FCN[12]、UNet[13]和 DeepLab v3+[14]等。目前也有不少研究将语义分割模型引入表面缺陷检测任务中，彭磊等[15]将 UNet 模型用于道路缺陷检测，郭亚萍等[16]将 SegNet 网络应用于工件表面缺陷检测中。除了通用的分割模型以外，针对条纹形的表面缺陷不易分割这一问题，ScratchNet[17]利用基于交叉最大池化模块的特征金字塔结构，结合多层信息精确地提取各个方向的边界特征，更关注边界细节的分割；再利用空间注意上采样模块，确保低分辨率到高分辨率特征传递的有效性，提高了检测准确度。实验证明基于分割的模型在表面缺陷检测任务中也有着不错的表现。

然而，以上这些算法都依赖于卷积神经网络(CNN)[18]，并在其基础上发展而来，但卷积神经网络更关注局部特征而

收稿日期：2022-01-05；**修回日期：**2022-03-10 **基金项目：**国家重点研发计划项目(2018YFB1307401)

作者简介：张乃雪(1997-)，女，四川德阳人，硕士研究生，主要研究方向为计算机视觉、深度学习；钟羽中(1987-)，女(通信作者)，湖南益阳人，博士，主要研究方向为计算机视觉、图像处理(zyzc122@163.com)；赵涛(1988-)，男，四川巴中人，副教授，博导，博士，主要研究方向为先进控制、感知与人工智能算法；佃松宜(1972-)，男，湖北荆州人，教授，博导，博士，主要研究方向为先进控制、感知与人工智能算法.

忽视了全局特征，因此基于 CNN 的检测算法对小尺度缺陷的检测存在不足。Transformer[19]则能够捕捉到较大范围内的特征信息，因此更关注全局特征。随着学者们将 Transformer 应用于各个领域[20, 21]，Carion 等[22]提出了基于 Transformer 的端到端的目标检测模型——DETR 模型。DETR 模型将目标检测任务转换为集合序列预测任务，通过简单的 CNN 提取特征，再利用基于 Transformer 的编码-解码结构进行并行地预测。DETR 模型的提出和应用，为解决目标检测任务提供了一种全新的思路。

在工业产品质量检测中，很难获取到大量经专家标注后的且种类丰富的缺陷样本，因此训练集样本的数量非常有限。同时，工业产品表面常常出现尺寸较小且不显著的缺陷，这对视觉缺陷检测是一个巨大的挑战。针对这些问题，本文提出了一种基于 Smooth-DETR 的产品表面小尺寸缺陷检测算法。考虑到工业产品表面通常以重复的结构性纹理作为背景，缺陷的存在打破了表面纹理的连续性，因此该算法利用 DETR 模型的全局特征学习能力，减少了对训练样本数量的需求，提升了缺陷的检出率。该算法还结合 Smooth-L1 损失函数[3]和 GIoU 损失函数[23]作为边框回归损失，以提升对小尺寸缺陷的回归精度和模型训练效率。另外，还使用匈牙利算法来得到分类损失和回归损失的最优匹配，提升算法整体的检测准确率。与基于 CNN 的端到端的表面缺陷检测算法相比，由于 Smooth-DETR 使用了更少的卷积层，计算复杂度更低、参数量更小。

1 Transformer 模型

Transformer 模型最初在自然语言处理(NLP)中被提出。该模型舍弃了用于提取特征的各种类型的卷积运算[24, 25]anchor-based 和非极大值抑制(NMS)等后处理方法，而是全部由注意力机制组成，最初用于解决机器翻译任务。Transformer 模型主要由多个基于多头注意力机制的编码器和解码器构成，如图 1 所示。每个编码器和解码器由多头注意力、前馈网络和层归一化等构成，其基本组成如图 2 所示。其中，多头注意力机制[19]为注意力层提供了多组由查询向量(Query)、关键向量(Key)和值向量(Value)组成的权重矩阵。假设含有 h 个多头，则多头注意力机制定义为

$$\begin{aligned}&\text{Attention}(Q,K,V)=\text{softmax}(\frac{QK^{T}}{\sqrt{d_k}})V\\&Q_i=QW_i^Q,K_i=KW_i^K,V_i=VW_i^V,i=1,\dots,h\\&\text{head}_i=\text{Attention}(Q_i,K_i,V_i)\\&\text{Multihead}=\text{Concat}(\text{head}_1,\dots,\text{head}_h)W^o\end{aligned} \tag{1}$$

其中，$\sqrt{d_k}$ 缩放因子；W_i^Q,W_i^K,W_i^V,W^o 均为权重矩阵。经过训练后，输入向量被投影到不同的子空间中，使得模型可以关注不同位置的信息，提升了注意力层的性能。与基于 CNN 的特征提取器相比而言，由于 Transformer 包含了多个子空间，故更容易关注全局特征。

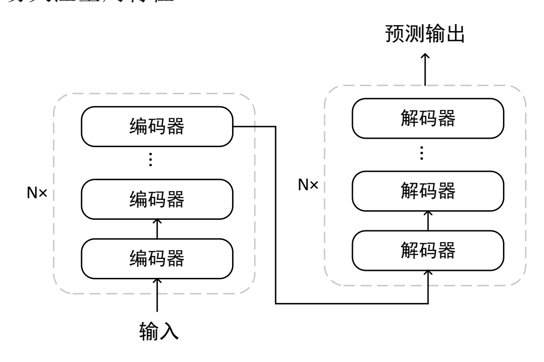


图 1 Transformer 模型结构

Fig. 1 Structure of Transformer

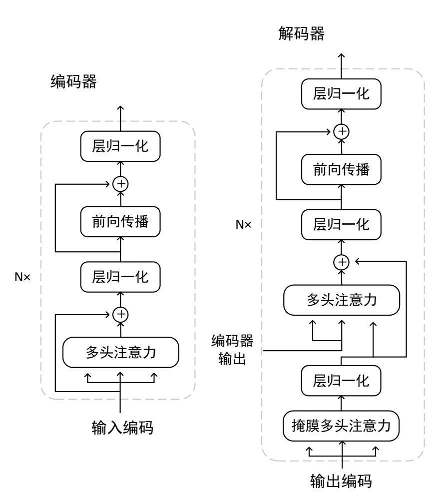


图 2 编码器和解码器的基本结构

Fig. 2 Structures of encoder and decoder

2 Smooth-DETR 算法

2.1 DETR 模型

DETR 为端到端的目标检测算法提出了一种全新的思路，它将 CNN 和 Transformer 模型相结合，并行地预测包含目标和背景在内的 N 个对象的类别信息。借助于 Transformer 模型更关注全局特征这一特性，DETR 模型也具有强大的全局特征学习能力。如图 3 所示，首先将输入图像分为大小相同的图像块，经过 CNN 网络对图像块进行特征提取，再利用 1×1 的卷积将特征图压缩成多个一维向量，和位置信息向量一并送入基于 Transformer 的编码器和解码器，N 个对象被转换成嵌入输出；最后，经过一个共享权重的前馈网络将这些嵌入输出向量并行地独立解码为 N 个类别和预测框。基于 Transformer 的编码-解码结构并行地预测整个输入，因此位置编码尤为重要。位置编码的计算如下：

$$\begin{aligned}PE_{(pos,2i)}&=\sin(pos/10000^{2i/d})\\PE_{(pos,2i+1)}&=\cos(pos/10000^{2i/d})\end{aligned} \tag{2}$$

其中，pos 表示图像块的位置，d 表示该向量维度；$2i$ 和 $2i+1$ 分别表示 d 中偶数维度和奇数维度。由三角函数性质可知，每个位置 $pos+k$ 都能用 pos 位置计算得到，且每个位置的所有维度都有独特的编码。

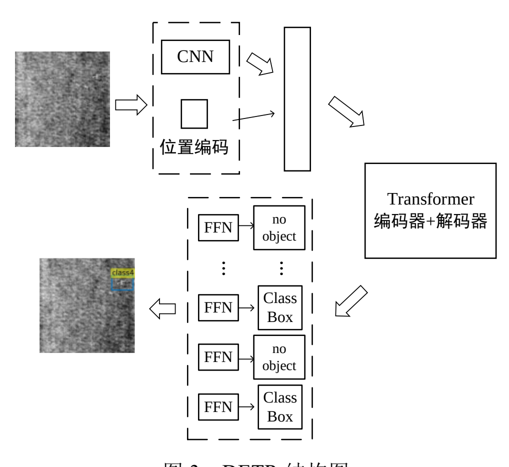


图 3 DETR 结构图

Fig. 3 Structure of DETR

假设输入图像大小为 H×W，图像通过 CNN 网络提取后的特征，再经过 1×1 的卷积降维，转换后得到 $D\times HW$ 大小的向量作为 Transformer 模型的输入。在基于 Transformer 的编码阶段，注意力矩阵大小为 (H×W)×(H×W)，如图 4 左图，表示了某一个 token 对应的注意力矩阵。注意力矩阵上的某一个点，实际对应此 token 所表示的特征块上两个不同的点，

如图 4 右图所示。又因为编码器的输入 token 数量与特征图像素个数相同，所以这就确定了一个框。因此，DETR 模型在目标检测任务具有独特的优势。此外，缺陷的存在破坏了产品表面纹理连续性，而全局特征学习能力强的 DETR 模型能挖掘到更丰富的表面纹理特征，因此其更易于实现产品表面的缺陷检测。

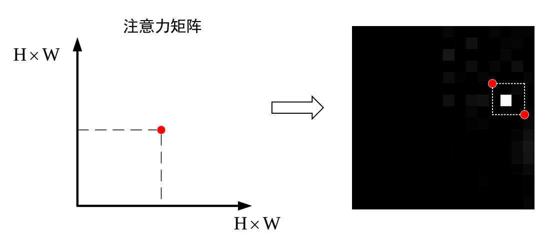


图 4　注意力矩阵与特征图对应关系

Fig. 4　Correspondence between attention matrix and feature map

此外，与基于 CNN 的目标检测不同的是，DETR 模型采用基于 Transformer 的编码器和解码器进行预测。因为减少了大量的卷积层，DETR 模型的计算复杂度和参数量都非常小。

2.2　损失函数设计

DETR 模型对解码器的嵌入输出向量进行解码预测；为了优化模型，需要通过损失函数不断减小预测值与标签值之间的偏差使模型达到最优。为了提升检测准确率，本文提出了 Smooth-DETR 算法，用 Smooth-L1 和 GIoU 损失函数相结合作为回归损失，对检测边框进行预测回归。Smooth-DETR 算法不仅有利于提升对小尺度缺陷的回归精度，增加检测准确率，还有利于提高训练鲁棒性和训练效率。除了回归损失外，本文还使用分类损失预测缺陷类别。

2.2.1 分类损失

交叉熵损失函数[25]可以很好地描述预测输出和期望输出之间的距离，通过不断学习优化模型预测每个类别的概率与 one-hot 形式的标签类别之间的距离，达到正确分类的目的。假设概率分布 $\hat{p}_{\sigma(i)}$ 为预测输出，c_i 为期望输出，则交叉熵损失函数定义为

$$L_{BCE}\left(\hat{p}_{\sigma(i)},c_i\right)=-\sum_{i=0}^{N}\log\hat{p}_{\sigma(i)}\left(c_i\right) \tag{3}$$

2.2.2 回归损失

在许多产品质量检测应用场景下，难以收集到大量的含有标注的缺陷样本，并且缺陷的形态大小各异，具有多样性；另外，产品表面常常出现小尺寸缺陷，而现有的在小尺寸缺陷样本上表现较好的模型往往结构复杂，计算量大。为了提升对小尺寸缺陷的检测准确率，本文采用 Smooth-L1 损失函数和 GIoU 损失函数相结合的方式，使算法不仅能够对小尺寸的缺陷稳定回归，提高检测准确率，还可以快速收敛到更高的精度。

$$L_{Smooth-L1}\left(b_{\sigma(i)},\hat{b}_i\right)=\begin{cases}\sum_{i=0}^{N}0.5\left(b_{\sigma(i)}-\hat{b}_i\right)^2, if\ \left|\hat{b}_i-b_{\sigma(i)}\right|<1\\ \sum_{i=0}^{N}\left|b_{\sigma(i)}-\hat{b}_i\right|-0.5, other\end{cases} \tag{4}$$

其中，$b_{\sigma(i)}$ 表示第 i 个索引的期望框，$\hat{b}_i$ 为第 i 个索引的预测框。

相比于 L1 损失函数和 L2 损失函数而言，Smooth-L1 损失函数结合了二者的优点，其定义如式(4)。如图 5 所示，在训练初期，预测框和期望框的距离过大，Smooth-L1 损失函数很好地限制了预测框的梯度，避免了“梯度爆炸”，并且在保留了模型快速收敛特性的同时，使模型更加鲁棒；而在训练后期，预测框和期望框的距离过小，损失函数在 0 附近波动时也存在导数，模型可以收敛到更高精度。

但是，Smooth-L1 损失函数在求解预测框时，仅仅独立的使用了 4 个点的损失值，并没有关注到 4 个值之间的相关性[19]，不能真实地反映预测框和期望框之间的包含关系。因此，在计算回归损失时，还引入了 GIoU[19]，它将预测框当作一个整体进行回归。

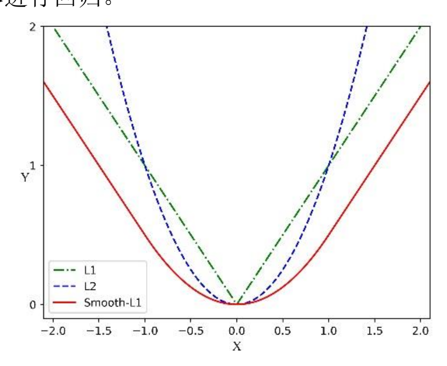


图 5　损失函数对比图

Fig. 5　Comparison of loss function

$$GIoU\left(b_{\sigma(i)},\hat{b}_i\right)=\frac{\left|b_{\sigma(i)}\cap\hat{b}_i\right|}{\left|b_{\sigma(i)}\cup\hat{b}_i\right|}-\frac{\left|B\left(b_{\sigma(i)},\hat{b}_i\right)\setminus\left(b_{\sigma(i)}\cup\hat{b}\right)_i\right|}{\left|B\left(b_{\sigma(i)},\hat{b}_i\right)\right|}$$

$$L_{GIoU}\left(b_{\sigma(i)},\hat{b}_i\right)=1-GIoU \tag{5}$$

式(5)展示是了 GIoU 的计算过程。其中 $B\left(b_{\sigma(i)},\hat{b}_i\right)$ 表示包围了 $b_{\sigma(i)}$ 和 $\hat{b}$ 的最小包围矩形；$B\left(b_{\sigma(i)},\hat{b}_i\right)\setminus\left(b_{\sigma(i)}\cup\hat{b}\right)$ 表示 $B(g)$ 中没有覆盖 $b_{\sigma(i)}$ 和 $\hat{b}$ 的面积。

除此之外，为了保证预测框与期望框之间的最大匹配，以及预测位置与类别一一对应，本文还采取了匈牙利算法[22]，即寻找增广路径，达成类别和预测框的最佳匹配。如式(6)所示，匈牙利算法选取损失值最小作为最佳匹配。

$$L_{Hungarian}\left(y,\hat{y}_{\sigma_i}\right)=\sum_{i=1}^{N}\left(L_{BCE}\left(\hat{p}_{\sigma(i)},c_i\right)+L_{box}\left(b_{\sigma(i)},\hat{b}_i\right)\right) \tag{6}$$

$$L_{box}=\alpha L_{GIoU}+\beta L_{Smooth-L1} \tag{7}$$

式(7)描述了预测框的损失函数，主要由 Smooth-L1 和 GIoU 共同决定，其中 α 和 β 分别表示 GIoU 损失函数和 Smooth-L1 损失函数的权重系数。基于此，本文通过 Smooth-L1 结合 GIoU 损失函数作为边框回归损失的策略，不仅提升了算法的检测准确率，使 Smooth-DETR 算法在小尺寸缺陷样本上也有很好的检测效果，还加快了算法收敛速度，提升了模型训练效率。

3　实验分析

3.1　数据集

为了证明本文所提算法对各种类型的缺陷，尤其是低对比度缺陷有着良好的检测效果，本文选取了两个公开数据集进行实验。

3.1.1 DAGM 2007

DAGM 2007 数据集是德国模式识别协会提供的，包含 10 种不同类型的人造缺陷数据集，如图 6 第一行和第三行所示，其中每个子类有约 80 张带缺陷的训练样本和 600 张测试集样本。在该数据集中，缺陷形态、大小各异；还包括了小尺寸缺陷。DAGM 2007 数据集中的所有样本大小均为 512×512，且在实验过程中均没有改变图像大小和数据增强。

3.1.2 KolektorSDD

KolektorSDD 数据集是 Kolektor 团队在受控工业环境下所采集的真实的电子换向器表面缺陷样本，如图 7(a)所示，共有 52 张缺陷图像和 347 张无缺陷图像。

3.2　实验环境

本文在两个公开数据集上进行了实验，并与现有的方法进行对比。所有的实验均在 Windows 系统下，GPU 采用的 NVIDIA GeForce RTX3060Ti，其显存为 16GB；CPU 采用的

Inter i7-10700F，内存为 32GB。本实验采用 Python 编写代码，深度学习框架为 PyTorch。

3.3 评价指标

为了定量地分析实验结果，验证本文所提方法的有效性，本文选取了 FP、FN、平均精确率(AP)、mAP、和 AUC 等作为评价指标[26]。FP 指假阳性，表示被误报为缺陷的正常样本；FN 指假阴性，表示被误报为正常样本的缺陷样本。AP 是指平滑后的 Precision-Recall 曲线与坐标轴所围成的面积，这使得 AP 能够精确地表示不同阈值下的综合模型性能；其中，Precision 和 Recall 分别表示查准率和查全率；mAP 则是所有类别 AP 的均值。AUC 是 ROC 曲线与坐标轴所围成的面积，ROC 曲线由 $FPR = FP/(TN+FP)$ 和 $TPR = TP/(TP+FN)$ 为横纵坐标构成，其中 TN 指预测正确的正常样本，TP 表示预测正确的缺陷样本。由于 ROC 曲线不随样本分布变换而变换，因此 ROC 曲线常常用于样本类别不均衡的情况。在 ROC 曲线不能直观地展示分类结果好坏时，往往选用 AUC 更清楚地描述分类结果好坏。

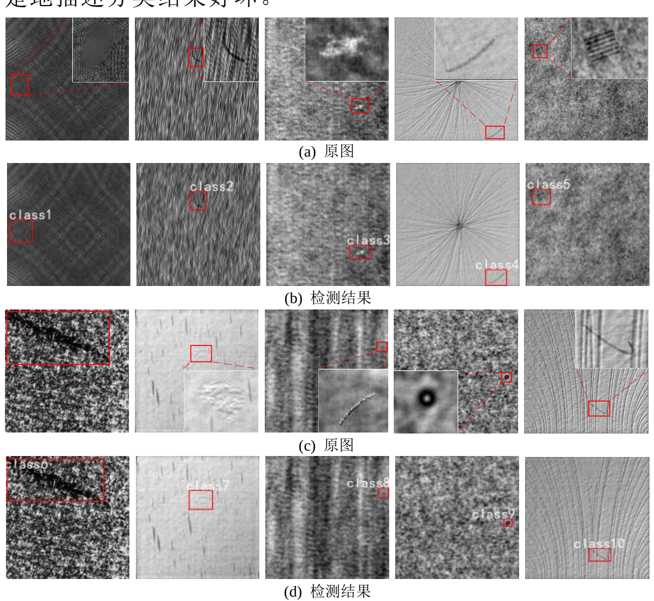


图 6 DAGM 2007 数据集的检测结果

Fig. 6 Detection results on DAGM 2007 dataset

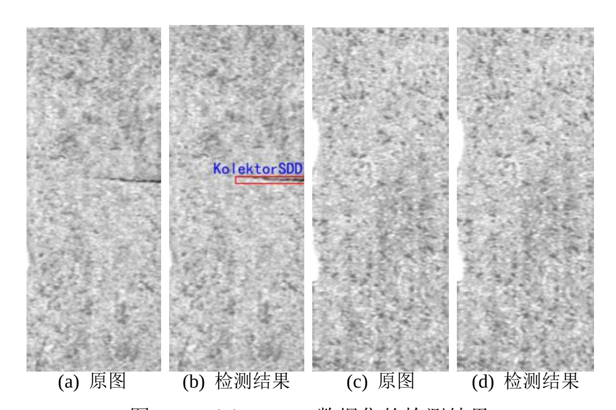


图 7 KolektorSDD 数据集的检测结果

Fig. 7 Detection results on kolektorsdd dataset

3.4 实验结果分析

3.4.1 与 DETR 模型实验对比分析

由于本文受 DETR 模型启发，同时为了验证本文所提的 Smooth-DETR 算法不仅在有限训练样本的情况下检测效果良好，而且在训练过程中可以快速收敛到较高精度。本文实验设置与 DETR 模型在 DAGM 2007 数据集上进行对比。实验中，分别在每个子类中选取 20 张缺陷图片作为训练集、5 张样本作为验证集；两种算法在整个训练过程中均选取 40 个 epoch，权重衰减系数为 0.0001，Batchsize 设置为 4。采取 AdamW 优化器，初始学习率为 0.0001，每 15 个 epoch 学习率降低 10 倍。

图 8 展示了两种算法在训练过程中不同的 epoch 时，训练集的错分率、总损失和 mAP 的变化。由于本文所提出的 Smooth-DETR 采用了 Smooth-L1 损失函数，相比于 DETR 模型，Smooth-L1 损失函数在预测框和期望框的距离十分相近时也存在导数，使模型可以收敛到更高精度，并且对小尺寸缺陷也能稳定回归；同时，因为 GIoU 损失函数在训练前期也具有梯度，结合 Smooth-L1 损失函数在训练前期快速收敛的优点，模型可以更加快速、稳定地收敛，提高了训练效率。另外，图 9 展示了验证集中各指标的变换，从图 9 的错分率变化曲线可以看出，大约在 14 个 epoch 时，Smooth-DETR 算法已经能够正确分类，并且整体精度高于 DETR 模型。

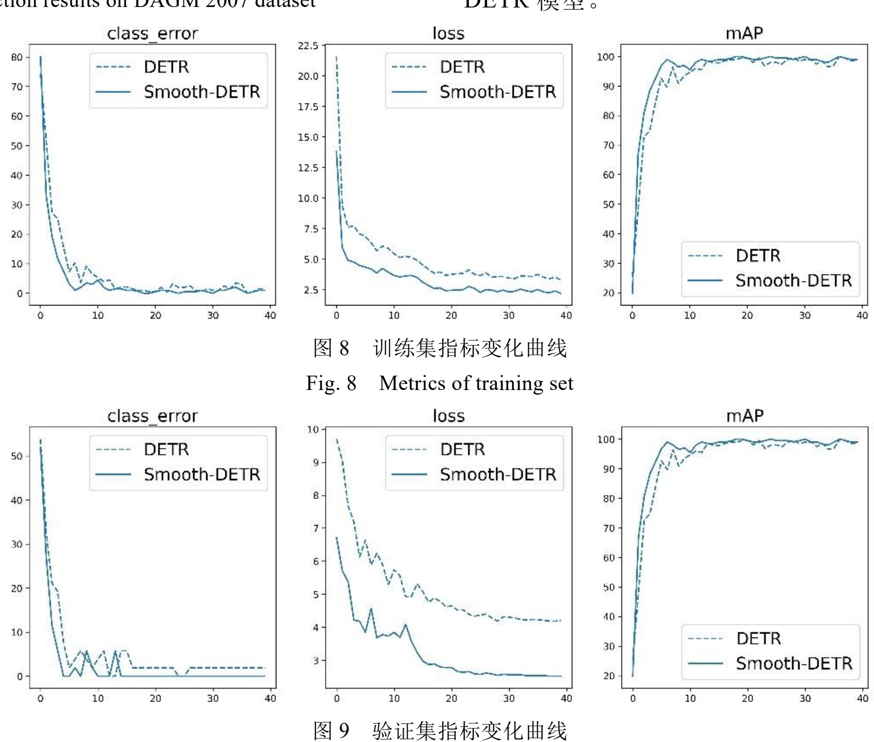


图 8 训练集指标变化曲线

Fig. 8 Metrics of training set

图 9 验证集指标变化曲线

Fig. 9 Metrics of validation set

为了证明 Smooth-DETR 在多种类型的缺陷上有更良好的表现，尤其对低对比度、小尺寸的缺陷检测效果对比 DETR 有所提升，本文选取了约 1 万张包含这 10 个不同的缺陷的样本作为测试集，进行实验；与 DETR 模型的检测结果对比如表 1 所示。

为了兼顾对错误样本的分类情况，并且分析算法对不同类型的缺陷的检测能力，选取了 AUC 指标对每个子类缺陷的检测结果进行定量地展示。表 1 对 DAGM 2007 数据集上每种缺陷的检测结果的 AUC 进行对比，从表中可以清晰地看出在 Class3、Class5 和 Class9 这些尺寸较小的低对比度缺陷上，由于 Smooth-L1 和 GIoU 损失函数相结合能对目标缺陷的边框更精确的回归，使得本文所提出的算法的检测性能有明显的提升。另外，Smooth-DETR 算法在 Class4 和 Class10 这类缺陷纹理与产品表面纹理具有相似性的样本上也有更好的表现。综合以上，说明了 Smooth-DETR 在各种形状和纹理的缺陷样本上有很好的检测效果，尤其对小尺寸缺陷也有着很好的检测效果。

表 1　AUC 指标对比结果

Tab. 1　Comparison results of AUC

数据集		AUC/%	
		Smooth-DETR	DETR[22]
DAGM 2007	Class 1	**97.6**	97.1
	Class 2	**98.7**	98.2
	Class 3	**98.1**	96.3
	Class 4	**98.3**	97.7
	Class 5	**100**	96.2
	Class 6	91.3	**95.5**
	Class 7	**100**	97.3
	Class 8	**95.8**	95.6
	Class 9	**100**	99.3
	Class 10	**100**	95.3

3.4.2 DAGM 2007 实验结果分析

为了直观地呈现 Smooth-DETR 算法在 DAGM 2007 数据集上的检测结果，图 6 展示了 10 种不同类型缺陷的检测结果。图 6 (a)和(c)是原图，图 6 (b)和(d)是对应的检测结果。为了清楚地展示缺陷，还在原图中添加了缺陷的放大图。在这 10 种不同类型的缺陷中，包括了 Class1、Class3、Class5 和 Class8 这类较小尺寸的低对比度缺陷，它们的缺陷纹理与产品表面纹理极为相似；以及 Class4 和 Class10 这类背景和缺陷都具有明显边缘特征的样本。另外，在该数据集中，大多类别的表面都是重复的结构性纹理，缺陷破坏了这种连续性，而本文所提出的算法具有强大的全局特征学习能力，因此本文所提出的算法在所有类型的缺陷上都有很好的检测效果，尤其是在 Class9 这类分辨率低于 17×17 像素的小尺寸缺陷上的检测结果依然很准确。从图 6 中可以看出，Smooth-DETR 不管在低对比度缺陷样本还是小尺寸样本上，检测框能够准确的回归缺陷所在位置，对不同的缺陷类型也能准确分类，有很好的检测效果。

进一步地，本文将所提 Smooth-DETR 方法与最新的表面缺陷检测算法——ScratchNet[17]、混合监督网络[11]、Faster-RCNN[27]、分类决策网络[28]和 YOLOv4-DefectSP[8]，在 DAGM 2007 数据集上进行比较，对比结果如表 2 所示。表 2 中 N 表示区域级标签样本，M 表示图像级标签样本。从表 2 中可以看出，Smooth-DETR 的检测性能要远高于 ScratchNet、Faster-RCNN、分类决策网络、YOLOv4-DefectSP 以及仅使用 5 张区域级标签样本和约 1000 张含图像级标签的样本训练的混合监督网络；其检测性能几乎媲美使用了有 15 张区域级标签样本和额外约 1000 张含图像级标签的样本训练的混合监督网络。表 2 证明了本文所提算法能适应多类产品表面缺陷检测任务，尤其是当训练数据不充分时。

表 2　DAGM 2007 数据集上的对比结果

Tab. 2　Comparison results on DAGM 2007 dataset

方法		AUC/%	mAP
ScratchNet[17](N=15, M=0)		93.7	94.8
混合监督网络[11]	(N=5, M=1000)	94.9	91.5
	(N=15, M=1000)	100	100
Faster-RCNN[27] (N=15, M=0)		88.3	90.4
分割决策网络[28] (N=15, M=0)		85.1	86.5
YOLOv4-DefectSP[8] (N=15, M=0)		89.8	92.2
Smooth-DETR(N=15, M=0)		98.0	98.5

在许多产品质量检测应用中，为了方便检测、提高检测效率以及借助于机器代替人工完成复杂的检测任务，检测设备逐步趋向于小型化便携式设备，不可避免地要求检测算法足够简单。也就是说，工业场景下的表面缺陷检测不仅需要关注检测结果，模型的大小和检测速度也同样重要。表 3 给出了以上六个模型的参数量和检测时间。如表 3 所示，ScratchNet 模型复杂度较高，会消耗大量的计算资源，检测时间也较长。虽然分割决策网络和 YOLOv4-DefectSP 的模型复杂度小、检测时间短，但是从表 2 中可以看出，这两个模型的检测性能相对较差。而对于通用的目标检测网络 Faster-RCNN 而言，因为使用了大量的卷积层，网络层数深，会消耗较多的检测时间，检测结果也并不理想。由于 Smooth-DETR 在特征提取阶段采用了基于 CNN 的网络，使得该算法具有一定的参数量。然而，正因为 Smooth-DETR 结合了 CNN 与 Transformer 的优点，该算法不仅有效地提取了全局特征还充分地利用了缺陷的局部特征。综上，Smooth-DETR 整体的实验结果最佳，保证检测速度的同时，还保持了较高的检测精确率。

表 3　参数量和检测时间对比

Tab. 3　Comparison of parameter and detection time

方法	参数量	检测时间/秒/张
ScratchNet[17]	51 Mio	0.92
混合监督网络[11]	30 Mio	0.67
Faster-RCNN[27]	41 Mio	1.03
分割决策网络[28]	2 Mio	0.32
YOLOv4-DefectSP[8]	10 Mio	0.41
Smooth-DETR	37 Mio	0.52

3.4.3 KolektorSDD 实验结果分析

为了证明 Smooth-DETR 算法适用于不同产品的表面缺陷检测，本文还在不同的数据集上进行实验。

图 7 展示了本文所提出的方法在真实的电子换向器缺陷 KolektorSDD 数据集上的检测结果。图 7 (a)和(c)分别是有缺陷原图和无缺陷原图；图 7 (b)和(d)分别是对应的检测结果图。从图 7 可以看出，Smooth-DETR 算法正确地识别电子换向器缺陷，说明 Smooth-DETR 算法除了在 DAGM 2007 数据集上有不错的检测效果，还可以适用于不同产品的检测。

目前，基于有监督的检测算法需要对大量的含有标签的样本进行长时间训练以获得较好的检测效果，而在实际产品质检应用中，很难获取到足够多的缺陷样本。本文将 Smooth-DETR 算法与 DeepLab v3+、UNet 和混合监督网络分别在 10 个和 20 个样本的训练集上进行了比较，结果如表 4 所示。DeepLab v3+和混合监督网络都是现有先进的表面缺陷检测算法。实验中，本文选用了性能表现最佳的超参数。值得说明的是，Smooth-DETR 算法仅需矩形框标注缺陷的大致范围，而 DeepLab v3、UNet 和混合监督网络需精确地标注缺陷像素。从表 4 可以看出，DeepLab v3+在 20 张训练集时，与 10 张训练集的检测精确率有明显的提升；UNet 对训练样本数量

敏感，这两个方法总体检测结果不佳。由于 Transformer 强大的全局特征学习能力，Smooth-DETR 算法在仅有 20 张样本进行训练后得到的模型，其 AUC 和平均精确率均高于 DeepLab v3+和混合监督网络。实验说明了 Smooth-DETR 中 Transformer 结构有更强的特征学习能力，可以从更少的样本中学习到缺陷特征，从而达到在少样本情况下的高检出率，用于解决质量检测应用中缺少大量的训练样本而造成的检测不准确的问题。

表 4　KolektorSDD 数据集上的对比结果

Tab. 4　Comparison results on kolektorsdd dataset

指标	DeepLab v3+[14]		UNet[13]		混合监督网络[11]		Smooth-DETR	
	10	20	10	20	10	20	10	20
AUC/%	77.5	95.7	88.1	81.8	**96.5**	98.8	95.6	**99.8**
AP/%	34.9	89.2	70.3	57.4	89.1	94.4	**91.0**	**98.9**
FP	38	2	6	8	3	2	**1**	**0**
FN	17	4	11	17	3	1	**4**	**1**

4　结束语

针对产品质量检测中，缺陷样本有限、形状不规则、尺寸较小且难以获取大量的缺陷样本用于训练检测模型等问题，本文提出了用于产品表面小尺寸缺陷检测的 Smooth-DETR 算法。该算法利用 DETR 模型强大的全局特征学习能力，提升了对破坏产品表面纹理连续性的缺陷的检出率；另外，该算法采用 Smooth-L1 和 GIoU 共同作为边框回归损失函数，提升了对小尺寸缺陷的检测准确率。同时该算法比现有的基于 CNN 的检测算法参数量更小，计算复杂度更低。实验结果显示，该算法在 11 种不同类型的缺陷数据集上都有不错的检测结果，说明了该算法具有普适性。与 DETR 算法相比，该算法不仅在训练阶段能更快速地回归到更高精度，还具有更高的平均检测精确率。此外，与现有的检测算法相比，Smooth-DETR 算法可以利用更少的训练样本，得到更好的检测性能。

参考文献：

[1] 任条娟，陈鹏，陈友荣，等．基于深度学习的多目标运动轨迹预测算法 [J]．计算机应用研究，2022，39 (1)：7．(Ren Tiaojuan, Chen Peng, Chen Yourong, *et al.* Multi-target motion trajectory prediction method based on deep learning [J]. Application Research of Computers, 2022, 39 (1): 7.)

[2] 赵玉卿，贾金露，公维军等．基于 pro-YOLOv4 的多尺度航拍图像目标检测算法 [J]．计算机应用研究，2021，38 (11)：3466-3471．(Zhao yuqing, Jia Jinlu, Gong Weijun, *et al.* Multi-scale aerial image target detection algorithm based on pro-YOLOv4 [J]. Application Research of Computers, 2021, 38 (11): 3466-3471.)

[3] Wei B, Hao K, Tang X, *et al.* Fabric defect detection based on faster RCNN [C]// international conference on artificial intelligence on textile and apparel. Springer, Cham, 2018: 45-51.

[4] Zhou X, Wang Y, Zhu Q, *et al.* A surface defect detection framework for glass bottle bottom using visual attention model and wavelet transform [J]. IEEE Transactions on Industrial Informatics, 2019, 16 (4): 2189-2201.

[5] Luo Q, Fang X, Liu L, *et al.* Automated visual defect detection for flat steel surface: A survey [J]. IEEE Trans on Instrumentation and Measurement, 2020, 69 (3): 626-644.

[6] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[7] He D, Xu K, Zhou P. Defect detection of hot rolled steels with a new object detection framework called classification priority network [J]. Computers & Industrial Engineering, 2019, 128: 290-297.

[8] Lian J, He J, Niu Y, *et al.* Fast and accurate detection of surface defect based on improved YOLOv4 [J]. Assembly Automation, 2021.

[9] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network [J]. arXiv preprint arXiv: 1503. 02531, 2015, 2 (7) .

[10] Zhou H, Jang B, Chen Y, *et al.* Exploring faster RCNN for fabric defect detection [C]// 2020 Third International Conference on Artificial Intelligence for Industries (AI4I) . IEEE, 2020: 52-55.

[11] Božič J, Tabernik D, Skočaj D. Mixed supervision for surface-defect detection: From weakly to fully supervised learning [J]. Computers in Industry, 2021, 129: 103459.

[12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]// International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[14] Chen L C, Zhu Y, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European conference on computer vision (ECCV) . 2018: 801-818.

[15] 彭磊，张辉．基于 U-net 的道路缺陷检测 [J]．计算机科学，2021，48 (S2)：616-619．(Peng Lei, Zhang Hui. Road defect detection based on U-net [J]. Computer Science, 2021, 48 (S2): 616-619.)

[16] 郭亚萍，顾智聪，彭宏京．SegNet 在工件表面缺陷检测中的应用 [J]．计算机工程与设计，2019，40 (10)：2979-2984．(Guo Yaping, Gu Zhicong, Peng Hongjing. The application of SegNet in the surface defect detection [J]. Computer Engineering and Design, 2019, 40 (10): 2979-2984.)

[17] Mei S, Cai Q, Gao Z, *et al.* Deep learning based automated inspection of weak microscratches in optical fiber connector end-face [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-10.

[18] 周飞燕，金林鹏，董军．卷积神经网络研究综述 [J]．计算机学报，2017，40 (06)：1229-1251．(Zhou Feiyan, Jin Linpeng, Dong Jun. Summary of convolutional neural network research [J]. Chinese Journal of Computers, 2017, 40 (06): 1229-1251.)

[19] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.

[20] 张晓旭，马志强，刘志强等．Transformer 在语音识别任务中的研究现状与展望 [J]．计算机科学与探索，2021，15 (9)：1578-1594．(Zhang Xiaoxu, Ma Zhiqiang, Liu Zhiqiang, *et al.* Research status and prospects of Transformer in speech recognition [J]. Journal of Frontiers of Computer Science & Technology, 2021, 15 (9): 1578-1594.)

[21] You J, Korhonen J. Transformer for image quality assessment [C]// 2021 IEEE International Conference on Image Processing (ICIP) . IEEE, 2021: 1389-1393.

[22] Carion N, Massa F, Synnaeve G, *et al.* End-to-end object detection with transformers [C]// European Conference on Computer Vision. Springer, Cham, 2020: 213-229.

[23] Rezatofighi H, Tsoi N, Gwak J Y, *et al.* Generalized intersection over union: A metric and a loss for bounding box regression [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 658-666.

[24] Dai J, Qi H, Xiong Y, *et al.* Deformable convolutional networks [C]// Proceedings of the IEEE international conference on computer vision. 2017: 764-773.

[25] Chollet F. Xception: Deep learning with depthwise separable

convolutions [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[26] Padilla R, Netto S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms [C]// 2020 international conference on systems, signals and image processing (IWSSIP) . IEEE, 2020: 237-242.

[27] Ren S, He K, Girshick R, *et al.* Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in neural information processing systems, 2015, 28.

[28] Tabernik D, Šela S, Skvarč J, *et al.* Segmentation-based deep-learning approach for surface-defect detection [J]. Journal of Intelligent Manufacturing, 2020, 31 (3): 759-776.